

Detection of Emerging Faults on Industrial Gas Turbines using Extended Gaussian Mixture Models

Yu Zhang^{1*}

School of Engineering, University of Lincoln,
Lincoln, LN6 7TS, U.K.
yzhang@lincoln.ac.uk

Chris Bingham¹

School of Engineering, University of Lincoln,
Lincoln, LN6 7TS, U.K.
cbingham@lincoln.ac.uk

Miguel Martínez-García¹

School of Engineering, University of Lincoln,
Lincoln, LN6 7TS, U.K.
mmartinez@lincoln.ac.uk

Darren Cox²

Siemens Industrial Turbomachinery LTD
Lincoln, LN5 7FD, U.K.
darren.cox@siemens.com

ABSTRACT

The paper extends traditional Gaussian Mixture Model (GMM) techniques to provide recognition of operational states and detection of emerging faults for industrial or complex systems. A Variational Bayesian (VB) method allows a GMM to cluster with its Mixture Components (MCs) to facilitate the extraction of steady-state operational behaviour — this is recognised as being a primary factor in reducing the susceptibility of alternative prognostic/diagnostic techniques which can initiate false-alarms resulting from control set-point

* Corresponding author: Tel: +44(0)1522 837938

and load changes. Furthermore, a GMM with an Outlier Component (GMMOC) is discussed and applied for direct fault detection. To demonstrate the efficacy of the proposed techniques, real-time measurements from operational Industrial Gas Turbines (IGTs) show that the resulting VBGMM facilitates the selection of the number of required MCs to cluster the data, and thereby provide essential input for operational signature recognition. Moreover, GMMOC is shown to facilitate the early detection of emerging faults. An advantage of the VBGMM over traditional pre-defined thresholds is the extraction of steady-state data during both full- and part-load cases, and a primary advantage of the GMMOC method is its applicability for novelty detection when there is a lack of prior knowledge of fault patterns. Results based on measurements taken from IGTs operating in the field are therefore also included which show that the techniques provide an integrated pre-processing, benchmarking and novelty/fault detection methodology.

Key Words: Novelty/fault detection, Gaussian mixture model, Variational Bayesian Gaussian mixture model, Gaussian mixture model with an outlier component.

1. INTRODUCTION

Industrial gas turbines (IGTs) are utilised globally with units generally acting as prime-movers for either pumps, generators or compressors, for both on- and off-shore systems. Various root-cause factors responsible for failures on such systems include vibration, shock, noise, heat, cold, dust, corrosion, humidity, rain, oil debris, flow, pressure and excessive operating speed [1]. A key challenge of condition monitoring in order to provide an 'early warning' of faults is to distinguish between sensor-based failure, component failure, and the normal operational transient behaviour of the

system e.g. due to control or load changes. With advances in instrumentation, communications hardware and computational capability [2], there has been an increased ability to realize such prognostic and diagnostic methods and provide remedial action and informed flexible maintenance scheduling prior to encountering unplanned downtime. This is especially pertinent as a result of the increasing operational speeds of IGTs [3].

Two popular categories of monitoring techniques have emerged over recent years viz. model-based and signal processing-based approaches [4]. Model-based approaches construct models (or virtual sensors) to estimate physical variables from which residuals are calculated and used as indicators of emerging failure modes [5]. However, to build an accurate dynamic model that can accommodate the full operating envelope of IGTs is, in general, computationally demanding. In such circumstances, direct signal processing and data fusion methods often provide for more practical and effective monitoring solutions [6]. It is this latter category of techniques that is considered here.

Traditional signal processing-based methodologies use techniques such as principal component analysis (PCA) [7][8], artificial neural networks (ANNs) [9] and data filter methods [10]. However, monitoring systems based on these algorithms are generally only applicable during steady-state operating conditions, since measurement transients caused by changes of loading or control action can generate ‘false alarms’. Many techniques are therefore only effective under very constrained operating regimes. For instance, [11] employed ANNs for fault detection on gas turbines during the engine

start-up phase, whilst [12] only considers solutions during steady-state operation. Typically, such techniques do not attempt to address the issue by incorporating implicit methods that discriminate between steady-state and transient behaviour as part of the fault detection system, and it is this aspect that is initially considered here [13].

By recognizing that steady-state measurement data is often superimposed by noise having a characteristic Gaussian distribution [14], i.e. contains practical ‘measurement transients’ that are not due to operational variations, the signals can be modelled through use of Gaussian Mixture Models (GMMs) [15]. This characterization of signals has previously been reported [16][17] and used in a related family of clustering methods based on GMMs, regarded as ‘soft clustering’ techniques, with the benefits for condition monitoring and fault detection explored in [16][17][18].

Here then, it is initially shown that GMMs provide a convenient mechanism to effectively discriminate between data relating to steady-state operation and that relating operation with transients, and is specifically regarded in this instance as a pre-processing tool for subsequent operational pattern discrimination [13]. An essential part of developing the GMM methodology is parameter fitting, which is often carried out using an expectation–maximization (EM) method [19]. However, this requires the number of the mixture components (MCs) in the model to be fixed a-priori [18]. Consequently, Bayesian-based frameworks are commonly used to provide a probabilistic inference on the data [20], and Variational Bayesian GMM techniques (VBGMMs) have been proposed [21] to provide improved performance [22][23] by automatically selecting the number of MCs in the GMMs. VBGMMs are able to classify

steady-state operation that can occur under full- or part-load conditions (e.g. 50% load). Additionally therefore, as well as identifying steady-state operation, the remaining data, including that associated with start-ups, shut-downs and load changing conditions, is also naturally separated and it can therefore be used as a pre-analysis tool for alternative dynamic scenarios, as reported in [24][25] for instance.

An additional property of GMMs that facilitates novelty/fault detection is the ability to filter novel class samples when the machine learning system has no a-priori knowledge [26]. Moreover, GMMs have been previously reported for the detection of ‘novel’ vibration signatures with experimental results showing good fault detection properties with known classifications [27]. However, the technique was sensitive to vibration characteristics that are not associated with the detection of wear or damage, and so remained susceptible to initiating false-alarms. As a consequence of these features, the outlier components are now included as background ‘noise’ for the GMM [28][29], and the resulting technique is considered as a GMM with an outlier component (GMMOC). When using GMMOC, measurements remain characterized by the GMM, but novel characteristics, i.e. measurements that have a very low probability of being clustered into existing distributions, are considered as outliers. It is the identification of such outliers that provides a robust mechanism for IGT fault detection considered here to provide an ‘early warning’ fault detection tool.

Key contributions of this paper are summarized as follows:

- 1) An extension to the use of GMMs, including VBGMM and GMMOC, is proposed for novelty/fault detection on industrial systems.

- 2) Steady-state operation is discriminated from transient operation using VBGMM.
- 3) GMMOC is used to indicate the presence of outliers, and hence the emergence of faults.
- 4) The efficacy of the proposed techniques is demonstrated from case studies (CSs) on IGTs, where CS1 is considered as a feasibility study of VBGMM and GMMOC for novelty detection, and CS2 demonstrates a real bearing fault case study.

2. METHODOLOGY

The stages in the proposed methodology are depicted in Table 1. To provide an application focus for the development of the algorithms, measurements taken from bearing vibration probes on IGTs during commissioning are used as an illustrative example. Operational pattern separation is achieved through the use of VBGMM, where the steady state data are distinguished for further analysis in this paper. Datasets from the identified transient operation can also be used for fault detection through start-up analysis, shut-down analysis and during load changes [6][24][25], which is not included in the current paper. The most relevant features are then extracted from the steady-state data and a statistical ‘fingerprint’ for the extracted features is obtained through the application of GMMOC.

Table 1. Stages of the Proposed Methodology

Step	Process	Purpose	Output
1	Sensor measurement collection (sampled or batch)	-	Collate data for analysis
2	VBGMM	Pre-process data to remove transient measurements	Measurements only of steady state operation
3	Statistical analysis	Extraction of important features within the data	Features for identifying emerging faults
4	GMMOC	Novelty detection of emerging faults	Early warning of emerging or imminent fault conditions

2.1 Underlying principles of GMM

The empirical probability distribution of sampled data can be estimated by a GMM using a linear combination of Gaussian distributions $N(x|\mu, \sigma)$ [15], e.g. as a sum of K Gaussian distributions with mean μ_k and standard deviation σ_k . The GMM with $K \times$ MCs is expressed as:

$$p(x|\theta, \mu, \sigma) = \sum_{k=1}^K \theta_k N(x|\mu_k, \sigma_k), \quad (1)$$

where x is a multidimensional variable and θ_k are the mixing coefficients that need to be chosen.

Let $X \equiv \{x_m\}$ represent M data samples, $m=1,\dots,M$, where each sample consists of a multidimensional variable x_m . Provided that x_m are statistically independent, the probability function of X can be expressed as:

$$p(X | \theta, \mu, \sigma) = \prod_{m=1}^M \left(\sum_{k=1}^K \theta_k N(x_m | \mu_k, \sigma_k) \right), \quad (2)$$

which constitutes the likelihood function, and by taking the natural logarithm gives:

$$L(X | \theta) = \ln p(X | \theta, \mu, \sigma) = \sum_{m=1}^M \ln \left(\sum_{k=1}^K \theta_k N(x_m | \mu_k, \sigma_k) \right). \quad (3)$$

The mixture density is then:

$$\rho_{mk} = \theta_k N(x_m | \mu_k, \sigma_k), \quad (4)$$

where k indicates the MC index and m indicates the data sample index. The conditional probability is calculated as:

$$p_{mk} = \frac{\rho_{mk}}{\sum_{j=1}^K \rho_{mj}}, \quad (5)$$

for a selected component k on the given data sample x_m .

Evaluation of (3) typically necessitates an EM optimization procedure to maximize the log-likelihood function [18] from the maximization step (termed the M-step):

$$\frac{\partial L}{\partial \mu_k} = 0, \quad \frac{\partial L}{\partial \sigma_k} = 0, \text{ and } \frac{\partial L}{\partial \theta_k} = 0. \quad (6)$$

The unknowns in (3) are solved in the expectation step (E-step), from (6), as follows:

1) choose an initialization of $\theta_k^{(0)}$, $\mu_k^{(0)}$ and $\sigma_k^{(0)}$,

2) iteratively update p_{mk} , θ_k , μ_k and σ_k until convergence to a desired tolerance, using:

$$p_{mk}^{(i)} = \frac{\theta_k^{(i)} N(x_m | \mu_k^{(i)}, \sigma_k^{(i)})}{\sum_{k=1}^K \theta_k^{(i)} N(x_m | \mu_k^{(i)}, \sigma_k^{(i)})}, \mu_k^{(i+1)} = \frac{\sum_{m=1}^M p_{mk}^{(i)} x_m}{\sum_{m=1}^M p_{mk}^{(i)}},$$

$$\sigma_k^{(i+1)} = \sqrt{\frac{1}{M} \frac{\sum_{m=1}^M p_{mk}^{(i)} \|x_m - \mu_k^{(i+1)}\|^2}{\sum_{m=1}^M p_{mk}^{(i)}}}, \text{ and } \theta_k^{(i+1)} = \frac{1}{M} \sum_{m=1}^M p_{mk}^{(i)}, \quad (7)$$

where M is the dimensionality of the data. For the special case $k = 1$, the whole dataset belongs to only 1 cluster, and the problem reduces to that of a Gaussian distribution fit.

2.2 Extension to VBGMM

A Variational Bayesian (VB) method can be used to determine the required number of MCs. Specifically, $M \times K$ binary latent variables $z_{mk} \in \{0,1\}$ are used to indicate which MCs the data sample clusters into. When forming a GMM using classical methods, K is selected a-priori. However, when using a VB method, K is resolved from the solutions of $\{z_{mk}\}$, termed Z . The joint probability distribution function of all variables is therefore,

$$p(X, Z, \theta, \mu, \sigma) = p(X | Z, \mu, \sigma) p(Z | \theta) p(\theta) p(\mu | \sigma) p(\sigma). \quad (8)$$

A lower bound on $p(X | \theta) = p(X | Z, \mu, \sigma) p(Z | \theta)$ can be determined using the VB method reported in [21]. Let $\Psi \equiv \{Z, \mu, \sigma\}$, and the marginal likelihood is expressed as:

$$p(X | \theta) = \int p(X, \Psi | \theta) d\Psi. \quad (9)$$

Next, a variational distribution is introduced, i.e. $q(\Psi) \approx p(\Psi | \theta)$. Through the use of Kullback-Leibler (KL) divergence [22], i.e. $D_{KL}(q \| p)$, (9) becomes,

$$\ln p(X | \theta) = D_{KL}(q \| p) + L(q), \quad (10)$$

where $L(q)$ indicates the lower bound. Minimization of the KL divergence can be achieved by maximizing $L(q)$ by selecting appropriate q distributions. $q(\Psi)$ can be rewritten as the product of $q_i(\Psi_i)$, in terms of the subsets $\{\Psi_i\} = \{Z, \mu, \sigma\}$, giving,

$$q(Z, \mu, \sigma) = q_Z(Z)q_\mu(\mu)q_\sigma(\sigma) \quad (11)$$

The best distribution for each term, $q_i^*(\Psi_i)$, can be solved using,

$$q_i^*(\Psi_i) = \frac{\exp(\mathbb{E}_{j \neq i}[\ln p(X, \Psi)])}{\int \exp(\mathbb{E}_{j \neq i}[\ln p(X, \Psi)]) d\Psi_i}, \quad (12)$$

where $\mathbb{E}_{j \neq i}[\cdot]$ denotes the expectation of $q_j(\Psi_j)$ for all $j \neq i$.

Since $\{Z, \mu, \sigma\}$ are mutually coupled, these can be calculated using the E-step:

- 1) Initialize the parameters, which are normally set to be small, real and positive [30],
- 2) Calculate $q(\Psi)$ in (11) through use of (12), and
- 3) Iteratively update until the pre-defined tolerance is met.

After calculated $q(\Psi)$, $L(q)$ can be obtained. Maximization of $L(q)$, which minimizes $\ln p(X | \theta)$, is again achieved by using the EM procedure. The M-step for calculating θ can be derived from (7), and the optimised distributions q_i can be obtained through the E-step as mentioned above. For brevity, the reader is directed to [30] for a more in-depth discussion of the VB framework on GMMs.

2.3 Principles of GMMOC

GMMOC extends the original GMM [28][29] approach by adding an outlier component that is modelled by a uniform distribution. The hybrid GMMOC model is then written as:

$$p(x) = \sum_{k=1}^K \theta_k N(x | \mu_k, \sigma_k) + \left(1 - \sum_{k=1}^K \theta_k\right) U(x). \quad (13)$$

Since $U(x)$ is uniformly distributed, the EM procedure described previously can still be employed to solve for the required parameters. The outlier component is normally assumed to be small initially (e.g. 0.01, and therefore the initialization of the mixing coefficients in GMM satisfies $\sum_{k=1}^K \theta_k = 0.99$). In this case, there are $(K+1)$ clusters, including K Gaussian distributions and one uniform distribution – the outlier component.

Parameters μ_k , σ_k and θ_k can be estimated from (7), and if the probability of a data sample, belonging to any of the K Gaussian MCs, is smaller than a pre-defined threshold, it is clustered to the outlier component, and therefore indicates a warning of an emerging fault – or facilitate novelty detection.

2.4 Feature extraction

Feature extraction provides an essential tool for reducing the dimensionality of raw data whilst keeping informative features [25]. Many feature extraction techniques have been reported and successfully applied, including the use of the Fast Fourier transform (FFT) and Discrete Wavelet Transform (DWT), all involving elaborate time-frequency transforms [31][32]. However, the data used in the following studies are taken from IGT

units in the field with sampling rates in the order of minutes, which excludes the use of frequency domain based methods as they do not satisfy traditional sampling rate criteria for the measured variables. In this case, statistical features of the data in the time domain are used, e.g. the peak value, root mean square, crest factor, kurtosis, clearance factor, impulse factor, shape factor and skewness, etc., and the most informative features can be identified through optimization methods [33]. For this study, in order not to divert the focus from the use of extended GMMs, only the most basic statistical attributes are employed—the mean (which carries information about measurement equilibrium) and standard deviation (which carries information about signal power) are used here as a proof of concept, and also for practical reasons since it has been observed empirically that these features are sufficient in most cases.

3. APPLICATION CASE STUDIES

The proposed methods are applied to monitor the vibration characteristics of fluid-film inlet- and output-bearings which typically support the compressor rotors of sub-15MW IGTs, Figure 1. The thrust bearings and journal bearings have operating speeds in excess of ~10,000 rpm. Radial and axial positions are monitored using non-contact probes. Two experimental case studies are now considered, both of which adopt the procedure in Table 1. CS1 uses measurements taken over a relatively short time period (1-month) and demonstrates the effectiveness of VBGMM for operational state discrimination (steady-state/transient behaviour), feature extraction, and the initial set-up of a benchmarking ellipse using GMMOC; while CS2 considers the analysis of longer periods of

measurement data (12-month) and aims to show the efficacy of GMMs for identifying longer-term emerging faults.

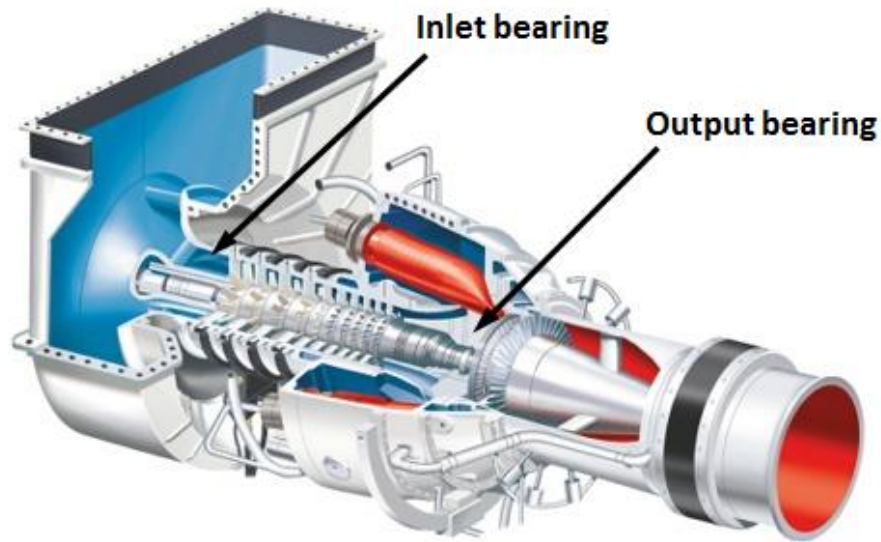


Figure 1. Locations of IGT bearing vibration sensors.

3.1 Case Study 1 : operation state discrimination, feature extraction and novelty detection

VBGMM is used to cluster measurements of output power (in terms of loading percentage) from a sub-15MW IGT to classify the unit's operational behaviour (Figure 2). One month (31 days) of daily data, each containing 1440 sample measurements, is used (i.e. sampling period = 1 minute). The resulting clusters from the power/load measurements are also shown in Figure 2 after applying VBGMM to each individual set of daily data. In line with the algorithm description, it can be seen that when the unit is considered to be operating normally, in steady-state, a classification label of 1 is

assigned, and that classification labels > 1 are assigned for all other detected cases
(cluster result given in

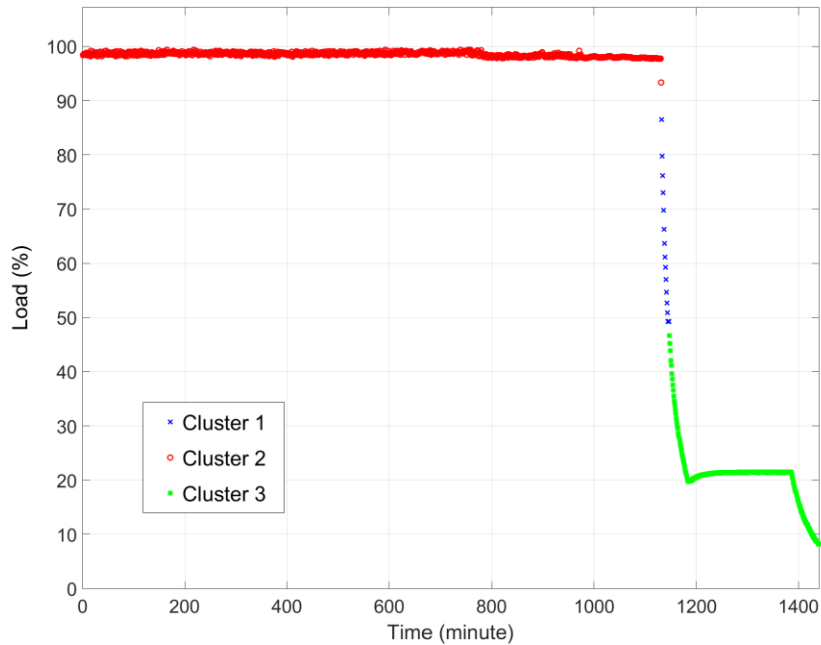


Figure 3). A classification label of 0 indicates constant null readings, and is precluded from further pattern analysis as the unit is considered to be shut down. Corresponding inlet bearing vibration measurements taken for the same 31 day period are shown in Figure 4. Having discriminated between transient and steady-state operation using

VBGMM, the steady-state data (days 1-12 and days 29-31) shown in

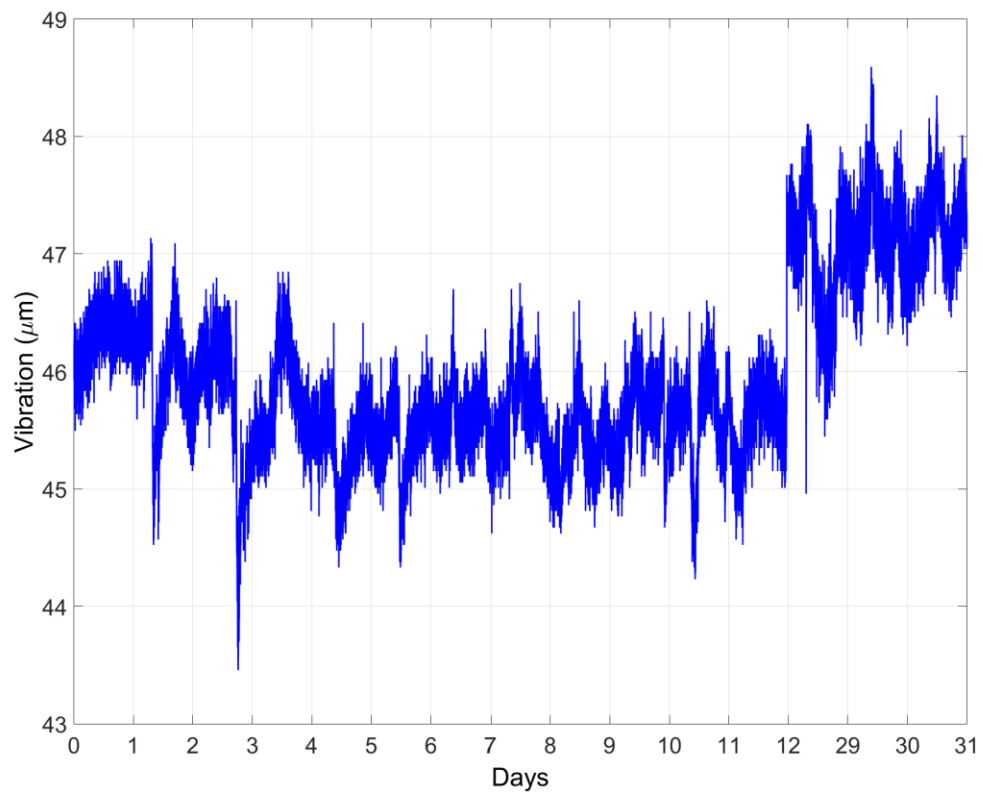


Figure 5 can be used for subsequent novelty detection.

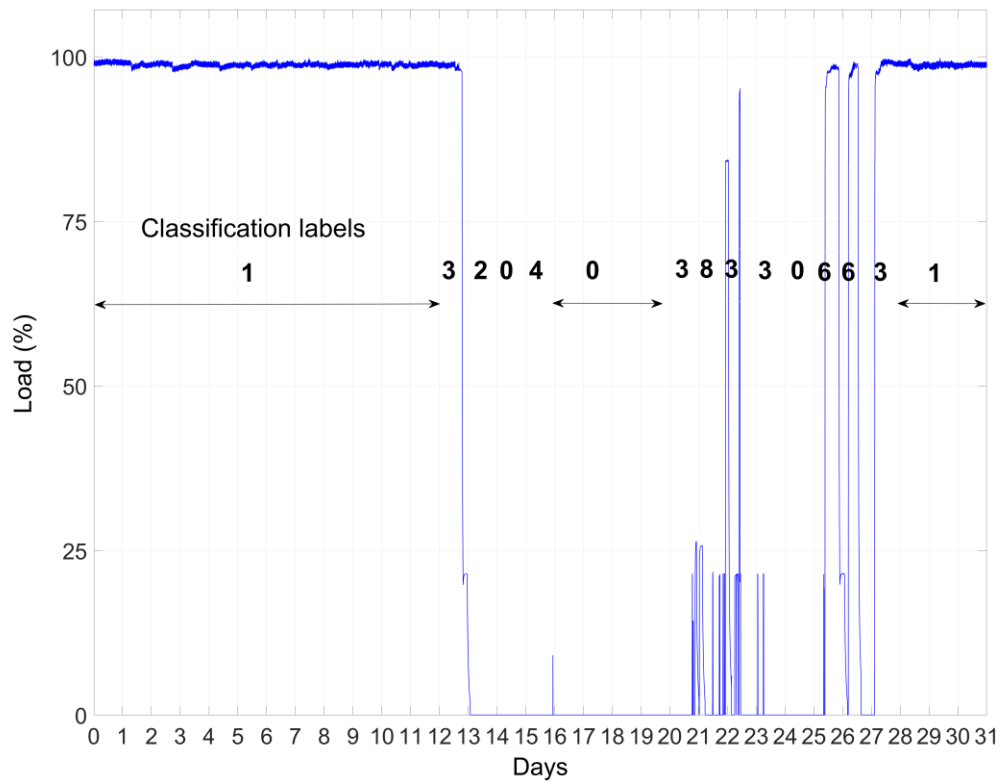


Figure 2. [CS1] Plot of load (%) vs. time (days): identification of clusters in power/load measurements (month 1—normal operation).

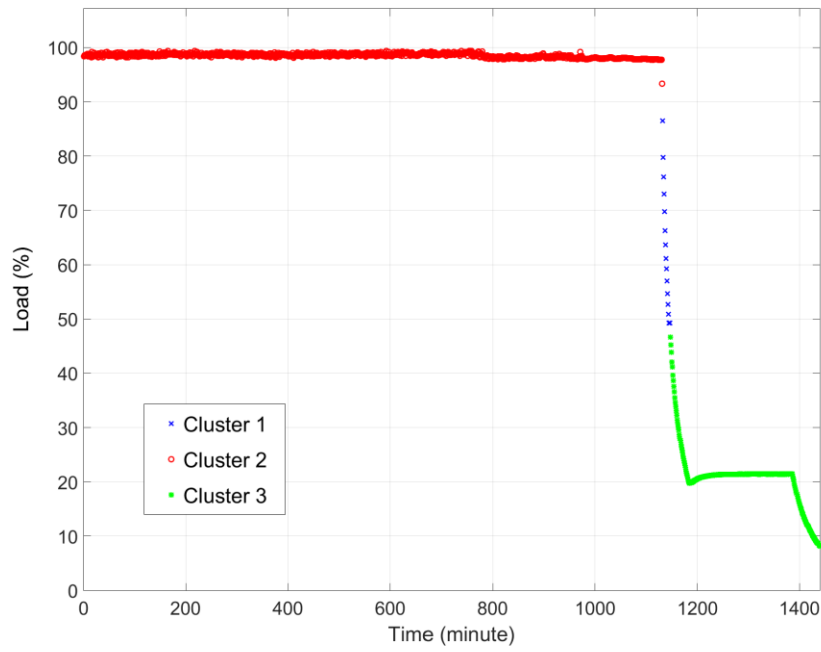


Figure 3. [CS1] VBGMM clustering result: Day 13 – 3 clusters identified, i.e. non-steady state data incorporating load changes.

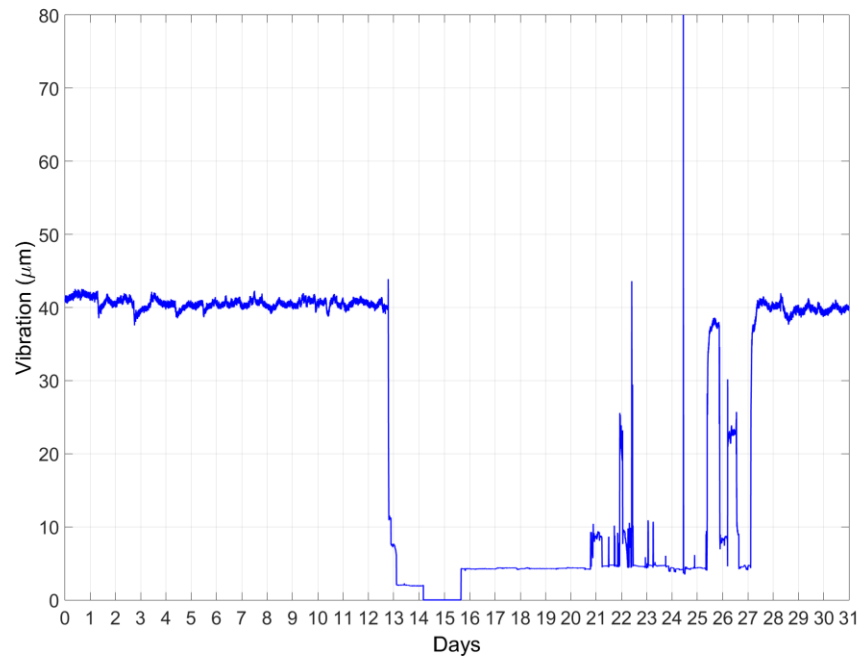


Figure 4. [CS1] Plot of vibration measurements vs. time (days) (month 1—normal operation).

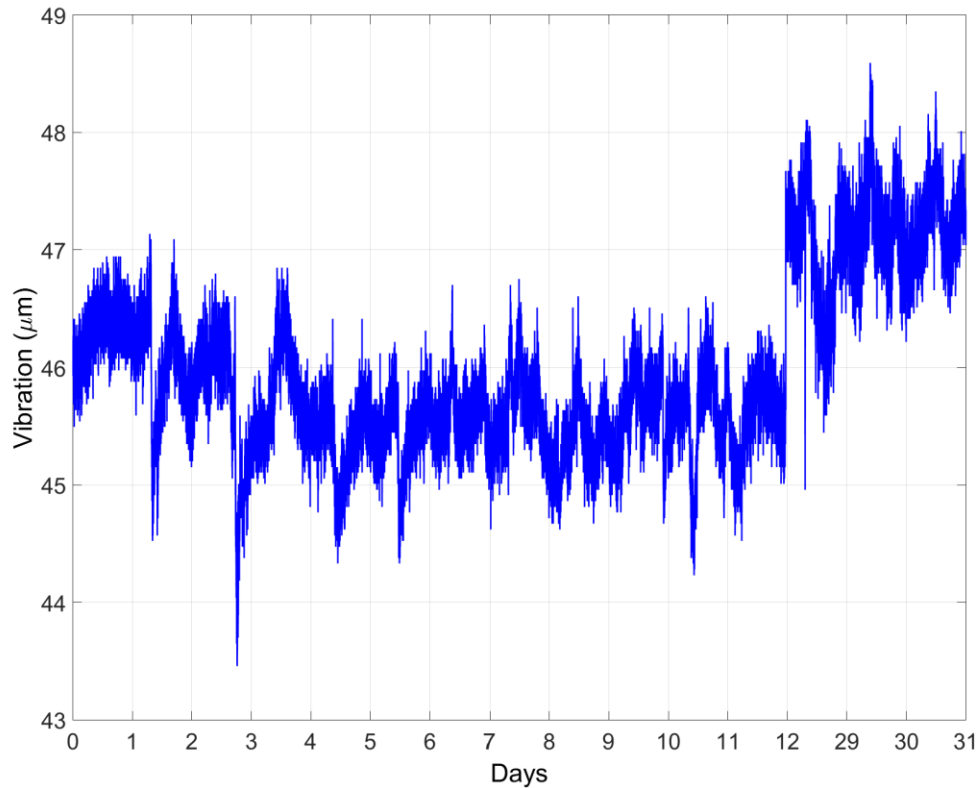


Figure 5. [CS1] Vibration data identified as steady state (month 1—normal operation).

Having identified appropriate data sets, feature extraction is used to capture important characteristics present in the data. The mean and the standard deviation of each day's data for Months 1—3 are calculated and given in Figure 6, to present a benchmarking envelope representing normal operation. Having effectively obtained an operational fingerprint of behaviour, measurements taken over subsequent months are then used and compared to the fingerprint.

Considering only the magnitude of vibration amplitudes, levels up to $\sim 50 \mu\text{m}$ are typically considered normal, with warnings at $\sim 70 \mu\text{m}$ and unit shutdown occurring at $\sim 90 \mu\text{m}$. Having obtained a fingerprint representing normal operation, GMMOC is applied to subsequent periods of data on a daily basis (in this case in Month 4, Figure 7).

It is well known that gradual bearing wear leading to failure is often preceded by gradual changes in vibration characteristics. Figure 8 provides an ellipse boundary drawn according to the 1-cluster GMM model (Eq.(13)). In general, the confidence level to identify outliers will be set according to application. For instance, in this case, a 99.99% confidence level does not discriminate the outliers (here, it indicates outliers in normal operation which may be caused by sensor malfunctioning), however, a 95% confidence level is clearly seen to be appropriate in this instance. The results from GMMOC are compared with the envelopes drawn from the original GMM, as shown in Figure 8, and the advantage of GMMOC over the original GMM is evident, since the outliers – days 9 and 10 are clearly identified even with a 99.99% confidence level of GMMOC in this case.

By considering the following month of measurement data, it is notable that the measurements have correctly been identified as outliers (day 9 and day 10 in this case), whereas measurements from days 1-4 are correctly considered to correspond to normal behaviour. Although it is not known at that stage if the increase in vibration in days 9 and 10 is related to a component fault, the measurements are considered as anomalies. Although CS1 is used as a proof of concept application of VBGMM and GMMOC for novelty detection, the bearing considered in the study did fail around 3 months after initially being identified using the extended GMM.

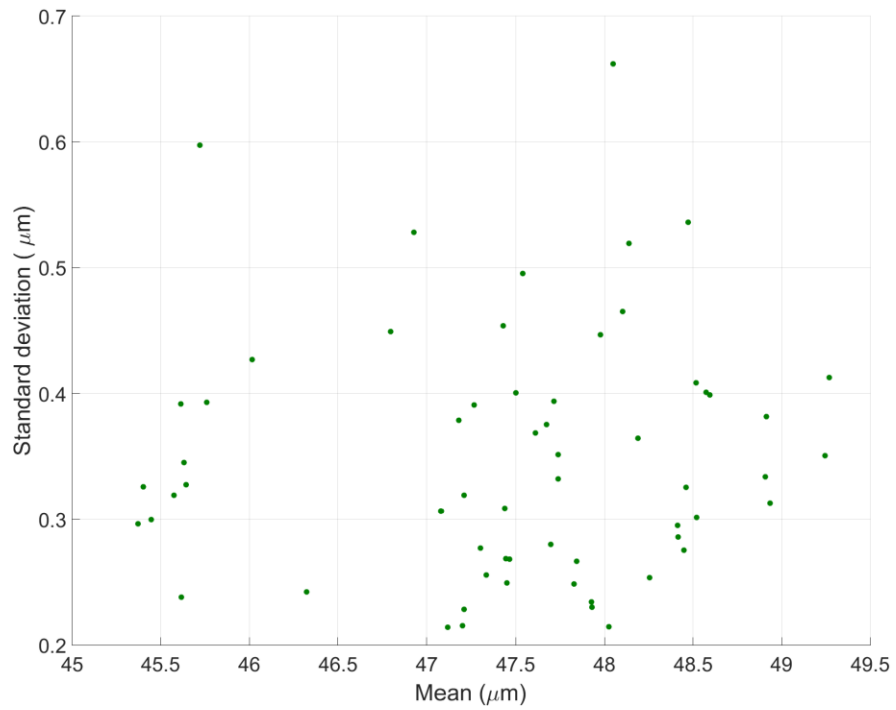


Figure 6. [CS1] Extracted features for 3 months' steady state data (months 1—3 – normal operation).

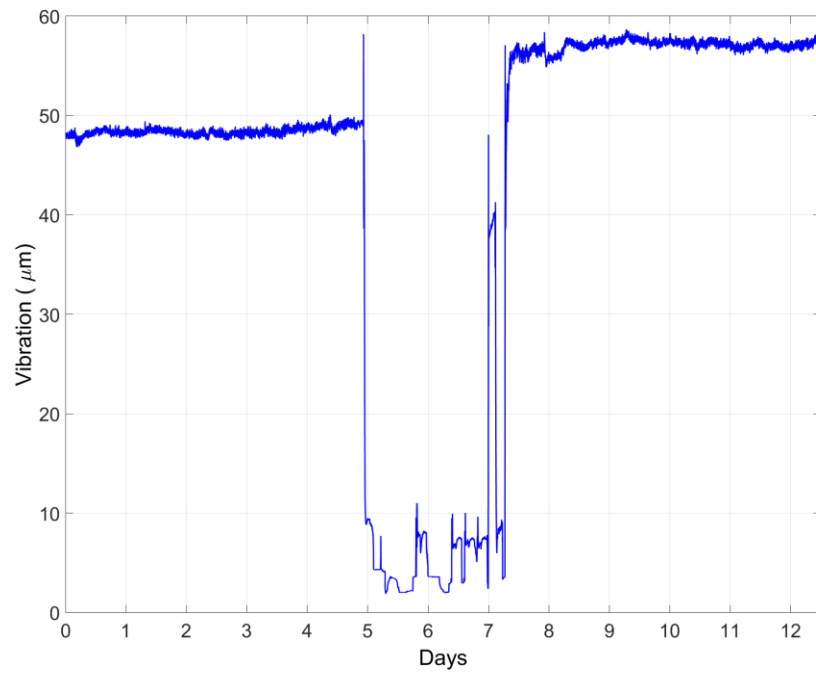


Figure 7. [CS1] Test data for novelty detection (in month 4).

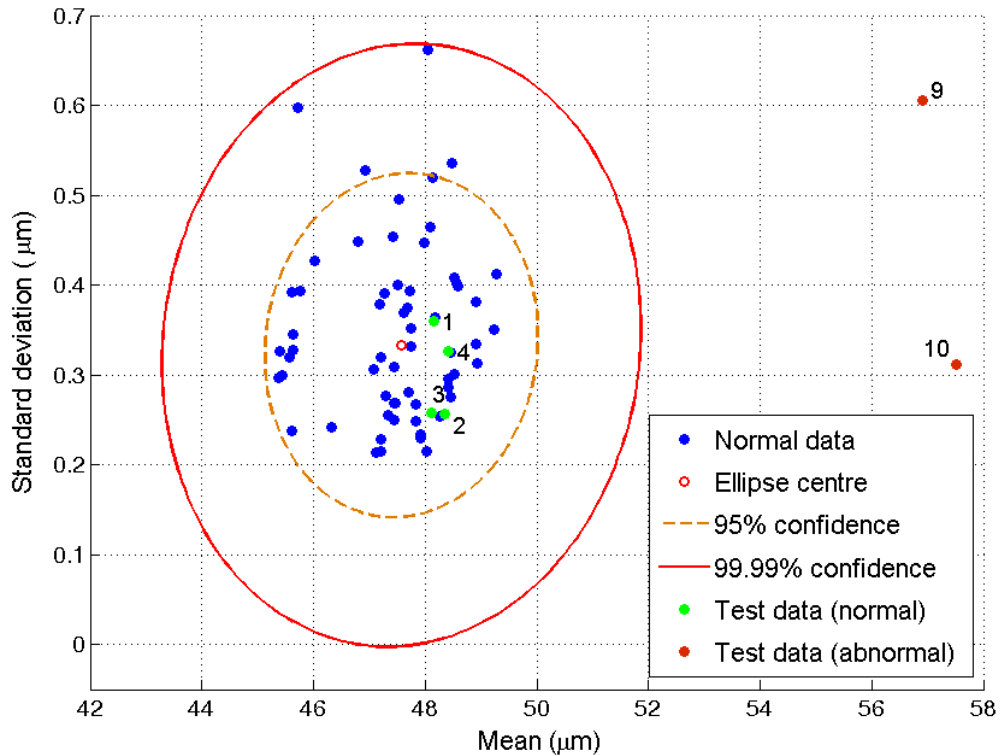


Figure 8. [CS1] GMMOC novelty detection results compared to those of the original GMM.

3.2 Case Study 2

CS2 uses measurements taken over a longer period of 12 months using a lower sampling rate; specifically 1 data sample taken every 9 minutes, as shown in Figure 9. Again the procedure depicted in Table 1 is applied.

Measurements from Months 1-4 are deemed to describe normal operational characteristics. VBGM is then applied to identify what is considered to be steady-state operational data for further analysis, **Figure 10**. Important characteristics are then determined — once again it is known from empirically studies that for this application case the mean and standard deviation are effective measures of underlying behaviour. Through the application of GMMOC and by clustering the extracted features into a single cluster, a fingerprint of normal operation is obtained, as shown in

Figure 11, where the 95% confidence envelope is considered as an early warning boundary in this instance, and the 99% confidence envelope as a fault detection boundary. Using measurements from Month 5 as testing data, it is shown in

Figure 11 that between days 20 to 25 of the 5th month, operation falls outside the 'normal fingerprint' ellipsoid, and is therefore identified as an outlier, thereby providing an early warning of expected failure, which was evidenced during the field service.

Referring again to measurements shown in Figure 9, on Day 140 of operation (Day 20 of Month 5) a transient in the vibration level is evident, and although the mean level ultimately returns to normal after Month 5, the variance continued to indicate evidence of emerging failure during and after Month 6, Figure 12. Considering Months 6 – 12 as a period of emerging fault, the number of fault patterns can be discriminated using VBGMM (2 clusters for faults in this case) and the fault pattern locations identified using GMMOC, as shown in Figure 12. It can be seen that Fault pattern 2 overlaps with the original fingerprint from Months 1–4, indicating in this case that ideally other feature extraction indices (e.g. those involved with loading conditions) could be used to further isolate this type of fault characteristic from that of normal operation. Between the 2 sets of fault patterns (Figure 12), some anomalies are apparent which are due to the increasing levels of vibration as the bearing deteriorates, such as the 20th to 25th days of the 5th month's operation.

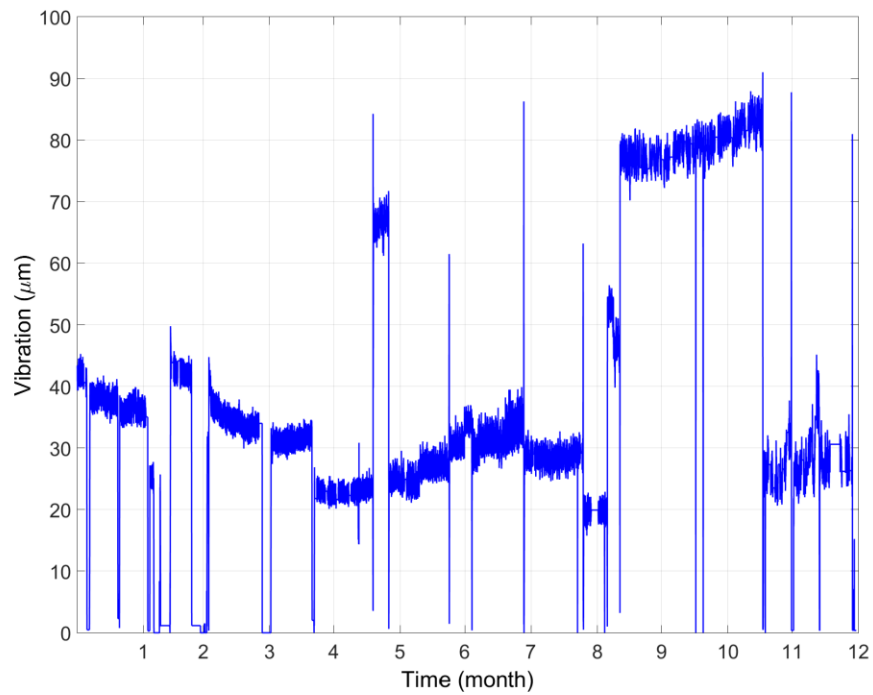


Figure 9. [CS2] IGT inlet bearing vibration information (months 1—12 – all operations including transients).

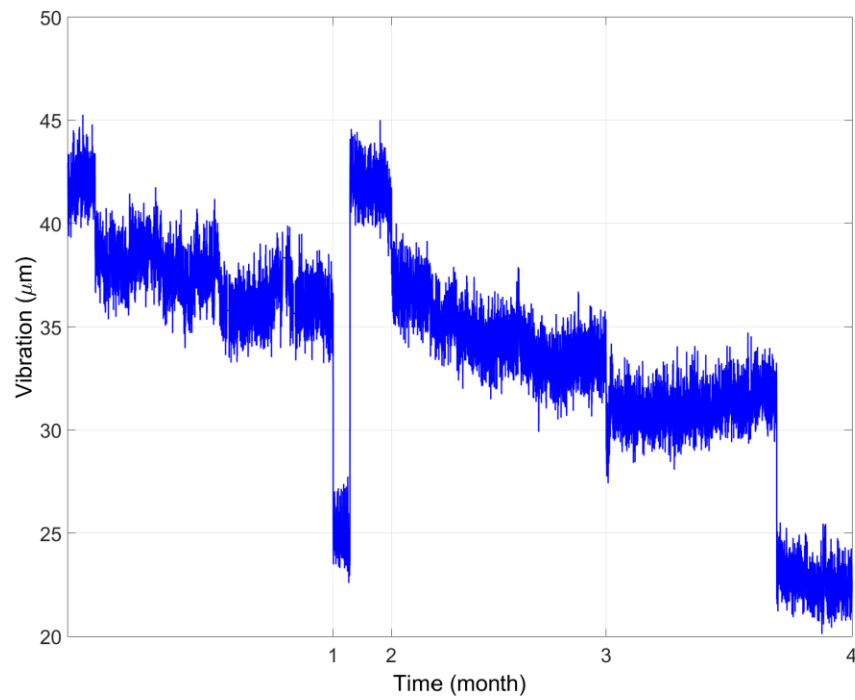


Figure 10. [CS2] Extracted steady-state vibration data for normal operation (months 1—4).

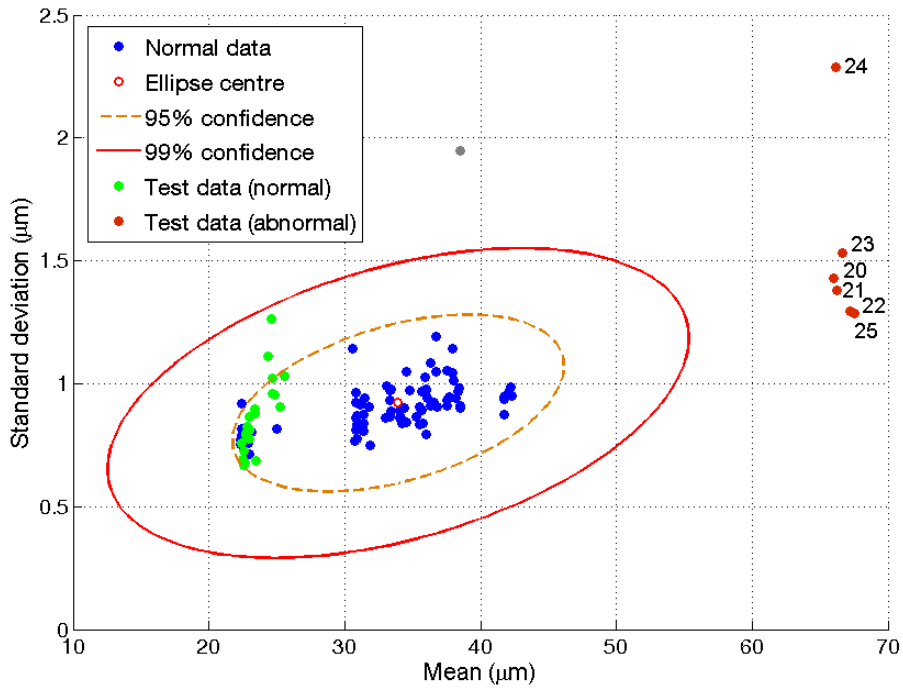


Figure 11. [CS2] Fingerprint envelope and novelty detection through GMMOC (months 1—4 and month 5).

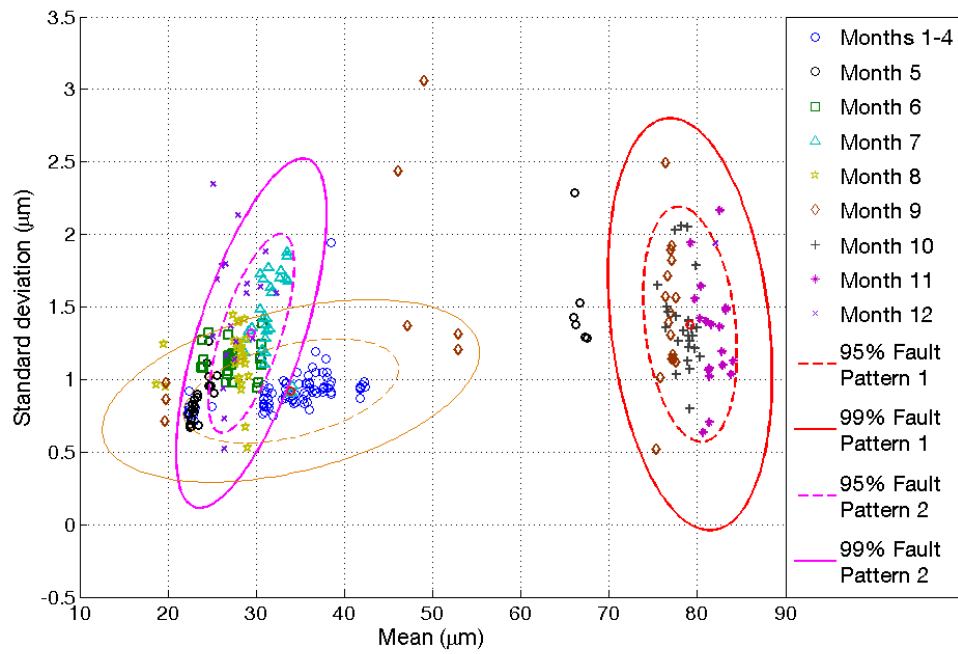


Figure 12. [CS2] Novelty detection and fault pattern location based on GMMOC (months 1—12).

In this instance the unit was shut-down for maintenance in order to prevent a catastrophic failure. The bearing was known to be undamaged in Month 1 as it was assessed in the previous service check (see Figure 13). Through subsequent decommissioning of the unit and investigation, vibration damage is evident on the inlet bearing, with excessive wear on the tilt pad shown in Figure 14. It can be clearly seen that there has been abnormal wear from the markings shown on both the shaft (Figure 15) and bearing pads (Figure 14). Through root cause analysis, the damage was attributed to an incorrectly specified lubricant oil cooler causing high temperatures in the lubricant oil.

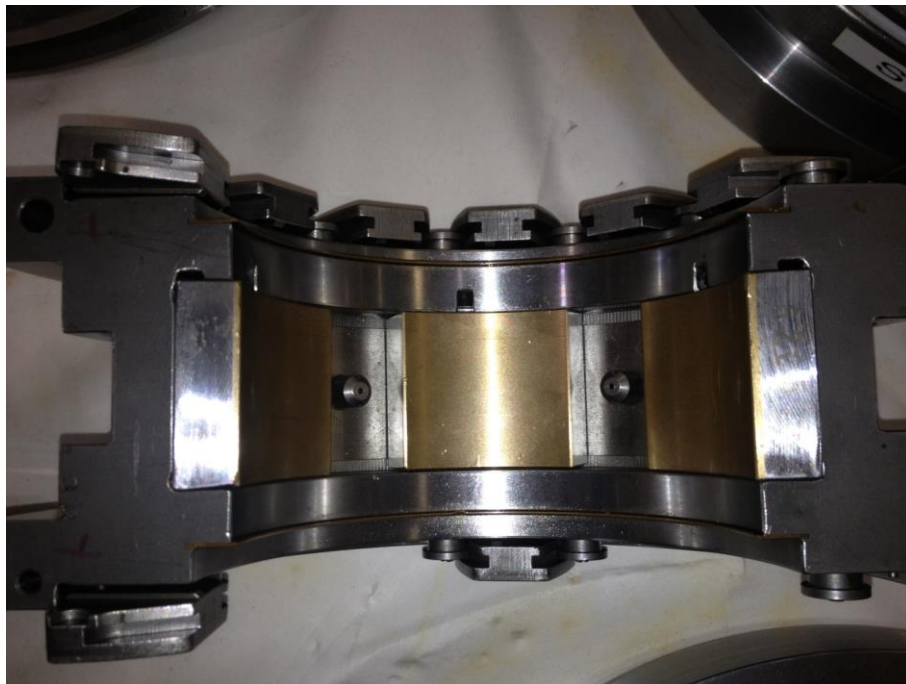


Figure 13. Normal IGT bearing with undamaged tilt pads.

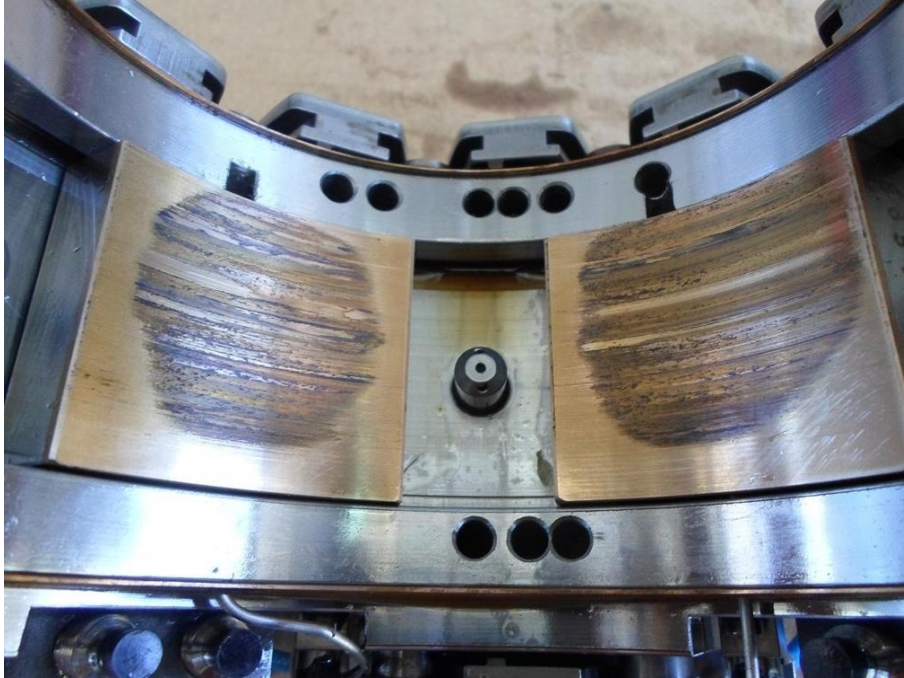


Figure 14. IGT inlet bearing showing vibration damage.

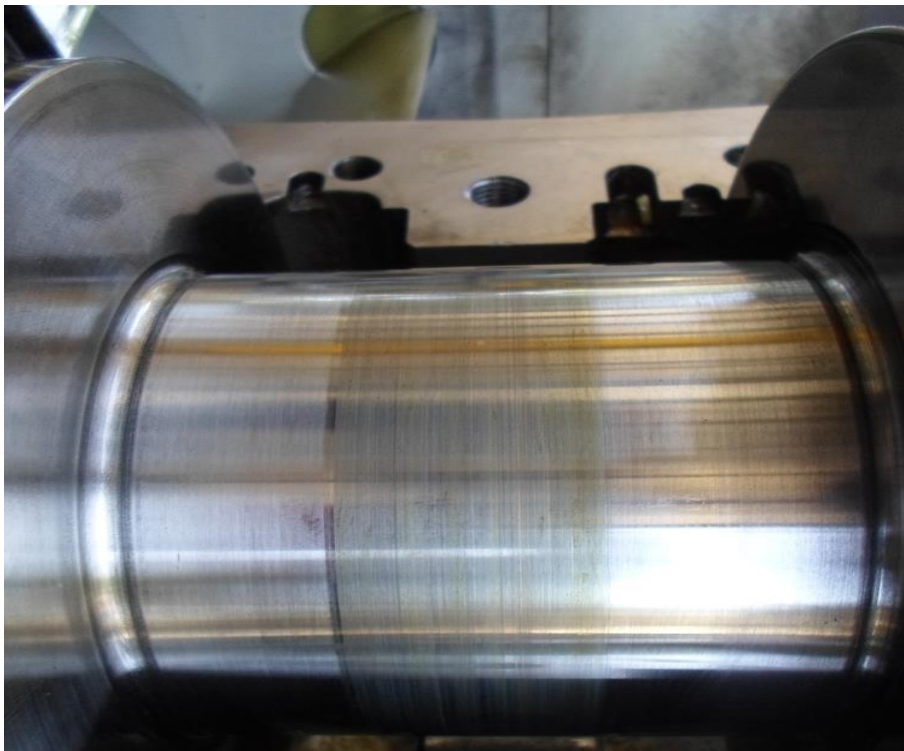


Figure 15. Compressor rotor inlet bearing shaft showing vibration damage.

4. CONCLUSION

The paper has developed and demonstrated extensions of GMMs to provide a highly practical pre-processing and novelty/fault detection tool. The main contributions of the paper are: 1) An automatic clustering method for VBGMM which identifies steady-state operational behaviour from transient operation, allowing the extraction of steady-state measurement segments for subsequent condition monitoring; and 2) A GMMOC method has been proposed and shown to provide a valuable tool for use as an early warning system of emerging failure through novelty detection. The presented techniques are currently being utilised in an industrial environment to monitor the operational status of a global fleet of IGTs. Although the experimental trials have focused on IGTs and bearing vibration measurements in this instance, the proposed methods are much more widely applicable to other industrial components and systems for pattern analysis, benchmarking and novelty/fault detection.

ACKNOWLEDGMENT

The authors would like to thank Siemens Industrial Turbomachinery, Lincoln, U.K., for providing research support and access to field data to support the research outcomes.

NOMENCLATURE

IGT	Industrial Gas turbine
GMM	Gaussian Mixture Model

VBGMM	Variational Bayesian Gaussian Mixture Model
GMMOC	Gaussian Mixture Model with an Outlier Component
PCA	Principal Component Analysis
ANN	Artificial Neural Network
MC	Mixture Component
EM	Expectation–Maximization
M-step	Maximization step
E-step	Expectation step
KL	Kullback-Leibler
CS	Case Study

REFERENCES

- [1] R. Kurz, K. Brun, M. Wollie, Degradation effects on industrial gas turbines, *J. Eng. Gas Turbines Power*, 131.6 (2009) 062401.
- [2] M.A. Zaidan, A.R. Mills, R.F. Harrison, P.J. Fleming, Gas turbine engine prognostics using Bayesian hierarchical models: A variational approach. *Mech. Syst. Signal Proc.*, 70-71 (2016) 120-140.
- [3] M. Maalouf, Gas turbine vibration monitoring, *Orbit Article*, 25.1 (2005) 1-7.
- [4] F. Lu, J. Huang, Y. Xing, Fault diagnostics for turbo-shaft engine sensors based on a simplified on-board model, *Sensors*, 12.8 (2012), 11061–11076.
- [5] H. Hanachi, J. Liu, A. Banerjee, Y. Chen, Sequential state estimation of nonlinear/non-Gaussian systems with stochastic input for turbine degradation estimation, *Mech. Syst. Signal Proc.*, 72-73 (2016) 32-45.
- [6] Y. Zhang, C. Bingham, Z. Yang, B.W.K. Ling, M. Gallimore, Machine fault detection by signal denoising—with application to industrial gas turbines, *Measurement*, 58 (2014) 230-240.

- [7] S. Yin, S.X. Ding, A. Haghani, H. Hao, P. Zhang, A comparison study of basic data-driven fault diagnosis and process monitoring methods on the benchmark Tennessee Eastman process, *J. of Process Control*, 22.9 (2012) 1567-1581.
- [8] Y. Zhang, C. Bingham, M. Gallimore, Fault detection and diagnosis based on extensions of PCA, *Advances in Military Technology*, 8.2 (2013): pp. 27-41.
- [9] C. Romesis, K. Mathioudakis, Setting up of a probabilistic neural network for sensor fault detection including operation with component Faults, *J. Eng. Gas Turbines Power*, 125.3 (2003) 634-641.
- [10] D.L. Simon, J.S. Litt, A data filter for identifying steady-state operating points in engine flight data for condition monitoring applications, *J. Eng. Gas Turbines Power*, 133.7 (2011) 603-610.
- [11] O. Uluyol, K. Kyusung, E.O. Nwadiogbu, Synergistic use of soft computing technologies for fault detection in gas turbine engines, *IEEE T. Systems Man Cybernetics, Part C: Applications and Reviews*, 36.4 (2006) 476-484.
- [12] Y. Zhang, C. Bingham, M. Gallimore, Applied sensor fault detection, identification and data reconstruction, *Advances in Military Technology*, 8.2 (2013) 13-26.
- [13] Y. Zhang, C. Bingham, M. Gallimore, J. Chen, Steady-state and transient operation discrimination by variational Bayesian Gaussian mixture models, *Proc. of IEEE Int. Workshop on Machine Learning for Signal Processing*, Southampton, U.K. 2013.
- [14] A. Jablonski, T. Barszcz, M. Bielecki, P. Breuhaus, Modeling of probability distribution functions for automatic threshold calculation in condition monitoring systems, *Measurement*, 46.1 (2013) 727-738.
- [15] Reynolds, D.A., *Gaussian Mixture Models*. Encyclopedia of Biometric Recognition. Springer, Heidelberg. 2008, pp. 659-663.
- [16] G. Yu, C. Li, J. Sun, Machine fault diagnosis based on Gaussian mixture model and its application, *Int. J. Adv. Manu. Tech.*, 48.4 (2010) 205-212.
- [17] J. Yu, Bearing performance degradation assessment using locality preserving projections and Gaussian mixture models, *Mech. Syst. Signal Proc.*, 25.7 (2011) 2573-2588.

- [18] C.M. Bishop, *Chapter 9: Mixture Models and EM. Pattern Recognition and Machine Learning. Series: Information Science and Statistics*. Springer, Heidelberg. 2006.
- [19] G. McLachlan, T. Krishnan, *The EM Algorithm and Extensions, 2nd Edition*. New York: Wiley, 2008.
- [20] A.R. Groves, Bayesian learning methods for modelling functional MRI, Ph.D. thesis, Department of Clinical Neurology, University of Oxford, U.K. 2010.
- [21] A. Corduneanu, C.M. Bishop, Variational Bayesian model selection for mixture distributions, *Artificial Intelligent Statistics*, 2001, pp. 27–34. Waltham, MA: Morgan Kaufmann.
- [22] K. Watanabe, S. Watanabe, Stochastic complexities of Gaussian mixtures in variational Bayesian approximation, *J. Machine Learning Research*, 7 (2006) 625-644.
- [23] Y. Zhang, C. Bingham, M. Gallimore, D. Cox, Novelty detection based on extensions of GMMs for industrial gas turbines, *Proc. of IEEE Int. Conf. on Computational Intelligence and Virtual Environments for Measurement Systems and Applications*, Shenzhen, China, 2015.
- [24] Y. Zhang, S. Cruz-Manzo, A. Latimer, Start-up vibration analysis for novelty detection on industrial gas turbines. *Proc. of XI Int. Symp. on Industrial Electronics*, Banja Luka, Bosnia and Herzegovina, 2016.
- [25] Y. Zhang, M. Gallimore, C. Bingham, J. Chen, Y. Xu, Hybrid hierarchical clustering – piecewise aggregate approximation, with applications, *Int. J. Comp. Intel. Appl.* 15, (2016) 1650019.
- [26] M. Markou, S. Singh, A neural network-based novelty detector for image sequence analysis, *IEEE T. Pattern Anal. Mach. Intell.*, 28.10 (2006) 1664-1677.
- [27] T. Heynsa, P.S. Heyns, J.P. DeVilliers, Combining synchronous averaging with a Gaussian mixture model novelty detection scheme for vibration-based condition monitoring of a gearbox, *Mech. Syst. Signal Proc.*, 32 (2012) 200-215.
- [28] M.K. Titsias, Unsupervised learning of multiple objects in images, Ph.D. thesis, School of Informatics, University of Edinburgh, U.K. 2005.
- [29] P. Coretto, The noise component in model-based clustering, Ph.D. thesis, Department of Statistical Science, University of London, U.K. 2008.

- [30] N. Nasios, A.G. Bors, Variational learning for Gaussian mixture models, IEEE T. System Man Cybernetics, Part B: Cybernetics, 36.4 (2006) 849-862.
- [31] P. Jayaswal, A. K. Wadhwani, K. B. Mulchandani, Machine Fault Signature Analysis, Int. J. of Rotating Machinery, 2008(2008) 583982.
- [32] S. Ding, H. Zhu, W. Jia, C. Su, A survey on feature extraction for pattern recognition, Artificial Intelligence Review, 37.3 (2012) 169-180.
- [33] Z. Huo, Y. Zhang, P. Francq, L. Shu, J. Huang, Incipient fault diagnosis of roller bearing using optimized wavelet transform based multi-speed vibration signatures, IEEE Access, (2017).